

A Faster Algorithm for Asymptotic Communication for Omniscience

Ni Ding*, Chung Chan[†], Qiaoqiao Zhou[‡], Rodney A. Kennedy* and Parastoo Sadeghi*

Abstract—We propose a modified decomposition algorithm (MDA) to solve the asymptotic communication for omniscience (CO) problem where the communication rates could be real or fractional. By starting with a lower estimation of the minimum sum-rate, the MDA algorithm iteratively updates the estimation by the optimizer of a Dilworth truncation problem until the minimum is reached with a corresponding optimal rate vector. We also propose a fusion method implementation of the coordinate-wise saturation capacity algorithm (CoordSatCapFus) for solving the Dilworth truncation problem, where the minimization is done over a fused user set with a cardinality smaller than the original one. We show that the MDA algorithm is less complex than the existing ones. In addition, we show that the non-asymptotic CO problem, where the communication rates are integral, can be solved by one more call of the CoordSatCapFus algorithm. By choosing a proper linear ordering of the user indices in the MDA algorithm, the optimal rate vector is also the one with the minimum weighted sum-rate.

I. INTRODUCTION

Communication for omniscience (CO) is a problem proposed in [1]. It is assumed that there is a group of users in the system and each of them observes a component of a discrete memoryless multiple source in private. The users can exchange their information over lossless broadcast channels so as to attain *omniscience*, the state that each user obtains the total information in the entire multiple source in the system. The CO problem in [1] is based on an asymptotic source model, where the communication rates could be real or fractional. Meanwhile, coded cooperative data exchange (CCDE) problem proposed in [2] can be considered a non-asymptotic CO problem where the communication rates are required to be integral. By incorporating the idea of packet-splitting, the CCDE problem can be easily extended to an asymptotic setting.

Determining a rate vector that achieves omniscience with the minimum sum-rate is a fundamental problem in CO. Although the non-asymptotic CO problem has been frequently studied in the literature, there still does not exist an efficient algorithm for the asymptotic setting. The reasons are explained as follows. The submodularity of the CO problem has been shown in [3]–[9]. By designating a sum-rate, a submodular

function minimization (SFM) algorithm can check whether the sum-rate is achievable for CO and/or return an achievable rate vector with the given sum-rate. Since the SFM algorithm completes in strongly polynomial time, the remaining problem is how to adapt the sum-rate to the minimum. This problem is not difficult for non-asymptotic setting since every adaptation should be integral. For example, the authors in [6], [7] proposed efficient adaptation algorithms for non-asymptotic CO problem, the complexity of which only grows logarithmically in the total amount of information in the system.

However, when considering the asymptotic setting, it is not clear how to choose the step size in each adaptation (Improper step sizes may result in an infinite loop). More specifically, even if we know that a sum-rate is over/below the optimum, it is not sure how much we should decrease/increase from the current estimation. On the other hand, the authors in [10] proposed a divide-and-conquer (DV) algorithm for the asymptotic setting by repetitively running a decomposition algorithm (DA) in [11]. The idea is to first find the fundamental partition [3], the one corresponds to the minimum sum-rate, and then iteratively break each non-singleton element into singletons so that each tuple in the optimal rate vector is determined. However, the DA algorithm is able to not only determine the fundamental partition but also return an optimal rate vector, which we will explain in this paper. Therefore, those further divisions of the fundamental partition in the DV algorithm are not necessary.

In this paper, we propose a modified decomposition algorithm (MDA) for solving the asymptotic CO problem based on the DA algorithm in [11]. The MDA algorithm starts with a lower estimation of the minimum sum-rate. In each iteration, the step size is determined based on the finest/minimum partition of a Dilworth truncation problem. We prove the optimality of the output rate vector and show that the estimation sequence converges monotonically upward to the minimum sum-rate. In addition, we propose a fusion method implementation of the coordinate-wise saturation capacity algorithm (CoordSatCapFus) for solving the Dilworth truncation problem. In the CoordSatCapFus algorithm, the SFM in each iteration is done over a fused user set with a cardinality smaller than the original one. We show that the MDA algorithm can reduce the cubic calls of SFM (in the DV algorithm) to quadratic calls of SFM. We do an experiment to show that the fusion method in the CoordSatCapFus algorithm contributes to a considerable reduction in computation complexity when the number of users grows. We also discuss how to solve the non-asymptotic CO problem by one more run of the CoordSatCapFus algorithm. Finally, we show how to choose a proper linear ordering to

*Ni Ding, Rodney A. Kennedy and Parastoo Sadeghi are with the Research School of Engineering, the Australian National University (email: {ni.ding, rodney.kennedy, parastoo.sadeghi}@anu.edu.au). Part of the work of Ni Ding (email: ni.ding@inc.cuhk.edu.hk) has been done when she was a junior research assistant at the Institute of Network Coding, Chinese University of Hong Kong, from Nov 23, 2015 to Feb 6, 2016.

[†]Chung Chan (email: cchan@inc.cuhk.edu.hk) and Qiaoqiao Zhou (email: zq115@ie.cuhk.edu.hk) are with the Institute of Network Coding, Chinese University of Hong Kong.

solve the minimum weighted sum-rate problem.

II. SYSTEM MODEL

Let V with $|V| > 1$ be the finite set that contains the indices of all users in the system. We call V the *ground set*. Let $Z_V = (Z_i : i \in V)$ be a vector of discrete random variables indexed by V . For each $i \in V$, user i can privately observe an n -sequence Z_i^n of the random source Z_i that is i.i.d. generated according to the joint distribution P_{Z_V} . We allow users exchange their sources directly so as to let all users $i \in V$ recover the source sequence Z_V^n . We consider both asymptotic and non-asymptotic models. In the asymptotic model, we will characterize the asymptotic behavior as the *block length* n goes to infinity. In non-asymptotic model, the communication rates are required to be integer-valued.

Let $\mathbf{r}_V = (r_i : i \in V)$ be a rate (vector). We call \mathbf{r}_V an achievable rate if omniscience is possible by letting users communicate with the rates designated by \mathbf{r}_V . Let r be the function associated with \mathbf{r}_V such that $r(X) = \sum_{i \in X} r_i, \forall X \subseteq V$ with the convention $r(\emptyset) = 0$. For $X, Y \subseteq V$, let $H(Z_X)$ be the amount of randomness in Z_X measured by Shannon entropy [12] and $H(Z_X|Z_Y) = H(Z_{X \cup Y}) - H(Z_Y)$ be the conditional entropy of Z_X given Z_Y . In the rest of this paper, we simplify the notation Z_X to X . It is shown in [1] that an achievable rate must satisfy the Slepian-Wolf constraints:

$$r(X) \geq H(X|V \setminus X), \quad \forall X \subset V. \quad (1)$$

The interpretation of the Slepian-Wolf constraint on X is: To achieve CO, the total amount of information sent from user set X should be at least complementary to total amount of information that is missing in user set $V \setminus X$. The set of all achievable rate vectors is

$$\mathcal{R}_{\text{CO}}(V) = \{\mathbf{r}_V \in \mathbb{R}^{|V|} : r(X) \geq H(X|V \setminus X), \forall X \subset V\}.$$

A. Asymptotic and No-asymptotic Models

In an asymptotic CO model, the minimum sum-rate can be determined by the following linear programming (LP)

$$R_{\text{ACO}}(V) = \min\{r(V) : \mathbf{r}_V \in \mathcal{R}_{\text{CO}}(V)\} \quad (2)$$

and the set of all optimal rates is

$$\mathcal{R}_{\text{ACO}}^*(V) = \{\mathbf{r}_V \in \mathcal{R}_{\text{CO}}(V) : r(V) = R_{\text{ACO}}(V)\}.$$

In a non-asymptotic CO model, $H(X) \in \mathbb{Z}_+$ for all $X \subseteq V$ and the minimum sum-rate can be determined by the integer linear programming (ILP) $R_{\text{NCO}}(V) = \min\{r(V) : \mathbf{r}_V \in \mathcal{R}_{\text{CO}}(V) \cap \mathbb{Z}^{|V|}\}$. The optimal rate set is $\mathcal{R}_{\text{NCO}}^*(V) = \{\mathbf{r}_V \in \mathcal{R}_{\text{CO}}(V) \cap \mathbb{Z}^{|V|} : r(V) = R_{\text{NCO}}(V)\}$.

B. Corresponding CCDE Systems

CCDE is an example of CO, where the asymptotic model corresponds to the CCDE system that allows packet-splitting, while the non-asymptotic model corresponds to the CCDE system that does not allow packet-splitting. In CCDE, Z_i is the packet set that is obtained by user i , where each packet W_j belongs to a field \mathbb{F}_q . The users are geographically close to each other so that they can transmit linear combinations of

their packet set via lossless wireless channels to help the others recover all packets in $Z_V = \cup_{i \in V} Z_i$. In this problem, the value of $H(X)$ can be obtained by counting the number of packets in Z_X , i.e., $H(X) = |Z_X|$ and $H(X|Y) = |Z_{X \cup Y}| - |Z_Y|$.

Example II.1. Let $V = \{1, \dots, 5\}$. Each user observes respectively

$$\begin{aligned} Z_1 &= (W_a, W_c, W_e, W_f), \\ Z_2 &= (W_a, W_d, W_h), \\ Z_3 &= (W_b, W_c, W_e, W_f, W_g, W_h), \\ Z_4 &= (W_a, W_c, W_f, W_g, W_h), \\ Z_5 &= (W_b, W_d, W_f), \end{aligned}$$

where W_j is an independent uniformly distributed random bit. The users exchange their private observations to achieve the omniscience of $Z_V = (W_a, \dots, W_h)$. In this system, $R_{\text{ACO}}(V) = \frac{11}{2}$ and $R_{\text{NCO}}(V) = 6$. $\mathbf{r}_V = (0, \frac{1}{2}, 2, \frac{5}{2}, \frac{1}{2})$ is an optimal rate in $\mathcal{R}_{\text{ACO}}^*(V)$ for asymptotic model, while $\mathbf{r}_V = (0, 1, 2, 3, 0)$ is the optimal rate in $\mathcal{R}_{\text{NCO}}^*(V)$ for non-asymptotic model. The method to implement rate $\mathbf{r}_V = (0, \frac{1}{2}, 2, \frac{5}{2}, \frac{1}{2})$ is to let users divide each packets into two chunks of equal length and transmit according to rate $(0, 1, 4, 5, 1)$ with each tuple denotes the number of packet chunks. $(0, \frac{1}{2}, 2, \frac{5}{2}, \frac{1}{2})$ and $\frac{11}{2}$ are the normalized rate and sum-rate, respectively.

III. PRELIMINARIES

In this section, we list some existing results derived previously in [3]–[9], [13]–[16] for CO.

A. Submodularity and Nonemptiness of Base Polyhedron

It is shown in [15], [16] that the entropy function H is the rank function of a polymatroid, i.e., it is (a) normalized: $H(\emptyset) = 0$; (b) monotonic: $H(X) \geq H(Y)$ for all $X, Y \subseteq V$ such that $Y \subseteq X$; (c) submodular:

$$H(X) + H(Y) \geq H(X \cap Y) + H(X \cup Y) \quad (3)$$

for all $X, Y \subseteq V$. For $\alpha \in \mathbb{R}_+$, define the set function f_α as

$$f_\alpha(X) = \begin{cases} H(X|V \setminus X) & X \subset V \\ \alpha & X = V \end{cases}.$$

Let $f_\alpha^\#(X) = f_\alpha(V) - f_\alpha(V \setminus X) = \alpha - f_\alpha(V \setminus X), \forall X \subseteq V$ be the *dual set function* of f_α . It is shown in [3], [9], [14] that $f_\alpha^\#$ is intersecting submodular, i.e., $f_\alpha^\#(X) + f_\alpha^\#(Y) \geq f_\alpha^\#(X \cap Y) + f_\alpha^\#(X \cup Y)$ for all $X, Y \subseteq V$ such that $X \cap Y \neq \emptyset$. The polyhedron and base polyhedron of $f_\alpha^\#$ are respectively

$$\begin{aligned} P(f_\alpha^\#, \leq) &= \{\mathbf{r}_V \in \mathbb{R}^{|V|} : r(X) \leq f_\alpha^\#(X), \forall X \subseteq V\}, \\ B(f_\alpha^\#, \leq) &= \{\mathbf{r}_V \in P(f_\alpha^\#, \leq) : r(V) = f_\alpha^\#(V)\}. \end{aligned}$$

It is shown in [8], [9], [13] that $B(f_\alpha^\#, \leq) = \{\mathbf{r}_V \in \mathcal{R}_{\text{CO}}(V) : r(V) = \alpha\}$, i.e., $B(f_\alpha^\#, \leq)$ denotes the set of all achievable rates with sum-rate equal to α , and $B(f_\alpha^\#, \leq) \neq \emptyset$ if and only if $\alpha \geq R_{\text{ACO}}(V)$. In addition, $B(f_{R_{\text{ACO}}(V)}^\#, \leq) = \mathcal{R}_{\text{ACO}}^*(V)$ and $B(f_{R_{\text{NCO}}(V)}^\#, \leq) \cap \mathbb{Z}^{|V|} = \mathcal{R}_{\text{NCO}}^*(V)$.

Algorithm 1: Modified Decomposition Algorithm (MDA)

input : the ground set V , an oracle that returns the value of $H(X)$ for a given $X \subseteq V$ and a linear ordering $\Phi = (\phi_1, \dots, \phi_{|V|})$

output: \mathbf{r}_V which is a rate vector in the base polyhedron $B(\hat{f}_{R_{ACO}(V)}^\#, \leq)$, \mathcal{P}^* which is the fundamental partition and α which equals to $R_{ACO}(V)$

- 1 initiate $\mathcal{P} \leftarrow \{\{i\} : i \in V\}$ and $\alpha \leftarrow \sum_{X \in \mathcal{P}} \frac{H(V) - H(X)}{|\mathcal{P}| - 1}$;
- 2 $(\mathbf{r}_V, \mathcal{P}^*) \leftarrow \text{CoordSatCapFus}(V, H, \alpha, \Phi)$;
- 3 **while** $\mathcal{P}^* \neq \mathcal{P}$ **do**
- 4 update $\mathcal{P} \leftarrow \mathcal{P}^*$ and $\alpha \leftarrow \sum_{X \in \mathcal{P}^*} \frac{H(V) - H(X)}{|\mathcal{P}^*| - 1}$;
- 5 $(\mathbf{r}_V, \mathcal{P}^*) \leftarrow \text{CoordSatCapFus}(V, H, \alpha, \Phi)$;
- 6 **end**
- 7 return $\mathbf{r}_V, \mathcal{P}^*$ and α ;

Denote $\Pi(V)$ the set that contains all possible partitions of V and $\Pi'(V) = \Pi(V) \setminus \{V\}$. For $\mathcal{P} \in \Pi(V)$, let $f_\alpha^\#[\mathcal{P}] = \sum_{X \in \mathcal{P}} f_\alpha^\#(X)$. The Dilworth truncation of $f_\alpha^\#$ is [17]

$$\hat{f}_\alpha^\#(X) = \min_{\mathcal{P} \in \Pi(X)} f_\alpha^\#[\mathcal{P}], \quad \forall X \subseteq V. \quad (4)$$

If $\alpha \geq R_{ACO}(V)$, $\hat{f}_\alpha^\#$ is submodular with $\hat{f}_\alpha^\#(V) = \alpha$ and $B(\hat{f}_\alpha^\#, \leq) = B(f_\alpha^\#, \leq)$ [13, Lemma IV.7].

B. Minimum Sum-rate and Fundamental Partition

The authors in [9], [13] show that

$$R_{ACO}(V) = \max_{\mathcal{P} \in \Pi'(V)} \sum_{X \in \mathcal{P}} \frac{H(V) - H(X)}{|\mathcal{P}| - 1} \quad (5)$$

and $R_{NCO}(V) = \lceil R_{ACO}(V) \rceil$. Meanwhile, in the studies on secrecy capacity in [3]–[5], it is shown that maximum secrecy capacity in V equals to the multivariate mutual information (MMI) $I(V)$, which has a dual relationship with $R_{ACO}(V)$: $R_{ACO}(V) = H(V) - I(V)$, and the finest/minimal maximizer of (5) is called the *fundamental partition* and denoted by \mathcal{P}^* .

IV. ALGORITHM

In this section, we propose a MDA algorithm, the modified version of the DA algorithm in [11], in Algorithm 1 for solving the asymptotic CO problem and show how to extend it to solve the non-asymptotic one. The MDA algorithm starts with α , a lower estimation of $R_{ACO}(V)$, and iteratively updates it by the minimal/finest minimizer of the Dilworth truncation problem $\hat{f}_\alpha^\# = \min_{\mathcal{P} \in \Pi(V)} f_\alpha^\#[\mathcal{P}]$ until it reaches the optimal one. The finest minimizer of the Dilworth truncation problem and a rate vector in the base polyhedron $B(\hat{f}_\alpha^\#, \leq)$ are determined by the CoordSatCapFus algorithm in Algorithm 2. The CoordSatCapFus algorithm is a fusion method to implement the coordinate-wise saturation capacity (CoordSatCap) algorithm that is proposed in [16] and adopted in [11] for the Dilworth truncation problem. We list the notations in Algorithms 1 and 2 below.

Let χ_X be the characteristic vector of the subset $X \subseteq V$. We shorten the notation $\chi_{\{i\}}$ to χ_i for a singleton subset of V . Let $\Phi = (\phi_1, \dots, \phi_{|V|})$ be a linear ordering of V . For example, $\Phi = (2, 3, 1, 4)$ is a linear ordering of $V = \{1, \dots, 4\}$. In

Algorithm 2: Coordinate-wise Saturation Capacity Algorithm by Fusion Method (CoordSatCapFus)

input : the ground set V , an oracle that returns the value of $H(X)$ for a given $X \subseteq V$, α which is an estimation of $R_{ACO}(V)$ and a linear ordering $\Phi = (\phi_1, \dots, \phi_{|V|})$

output: \mathbf{r}_V which is a rate vector in $B(\hat{f}_\alpha^\#, \leq)$ and \mathcal{P}^* which is the minimal/finest minimizer of $\min_{\mathcal{P} \in \Pi(V)} f_\alpha^\#[\mathcal{P}]$

- 1 $\mathbf{r}_V \leftarrow (\alpha - H(V))\chi_V$; // $\mathbf{r} \in P(f_\alpha^\#, \leq)$ by doing so.
- 2 initiate $r_{\phi_1} \leftarrow f_\alpha^\#(\{\phi_1\})$ and $\mathcal{P}^* \leftarrow \{\{\phi_1\}\}$;
- 3 **for** $i = 2$ **to** $|V|$ **do**
- 4 determine the saturation capacity

$$\hat{\xi} \leftarrow \min\{f_\alpha^\#(\{\phi_i\} \cup \tilde{U}) - r(\{\phi_i\} \cup \tilde{U}) : U \subseteq \mathcal{P}^*\}$$
 and the minimal/smallest minimizer U^* ;
- 5 $U_{\phi_i}^* \leftarrow U^* \cup \{\phi_i\}$;
- 6 $\mathbf{r}_V \leftarrow \mathbf{r}_V + \hat{\xi}\chi_{\phi_i}$;
- 7 /* merge/fuse all subsets in \mathcal{P}^* that intersect with $\tilde{U}_{\phi_i}^*$ into one subset

$$\tilde{U}_{\phi_i}^* \cup \mathcal{X} \quad */$$
- 8 $\mathcal{X} \leftarrow \{X \in \mathcal{P}^* : X \cap \tilde{U}_{\phi_i}^* \neq \emptyset\}$;
- 9 $\mathcal{P}^* \leftarrow (\mathcal{P}^* \setminus \mathcal{X}) \cup \{\tilde{U}_{\phi_i}^* \cup \mathcal{X}\}$;
- 10 **endfor**
- 11 return \mathbf{r}_V and \mathcal{P}^* ;

Section V, we will show that by choosing a proper linear ordering of V the output rate \mathbf{r}_V of Algorithm 1 also minimizes a weighted sum-rate objective function. For $U \subseteq \mathcal{P}$ where \mathcal{P} is some partition in $\Pi(V)$, denote $\tilde{U} = \cup_{X \in U} X$, i.e., U is a fusion of all the subsets in U into one subset of V . For example, for $U = \{\{1, 3\}, \{2, 4\}, \{5\}, \{6\}\} \subset \{\{1, 3\}, \{2, 4\}, \{5\}, \{6\}, \{7\}\} \in \Pi(\{1, \dots, 7\})$, we have $\tilde{U} = \{1, \dots, 6\}$. By using these notations, we propose the MDA algorithm for the asymptotic CO problem and show that they can be easily extended to solve the non-asymptotic CO problem as follows.

A. Asymptotic Model

The optimality of the MDA algorithm for the asymptotic setting is summarized in the following theorem with the proof in Appendix A, where every step in the CoordSatCapFus algorithm is explained.

Theorem IV.1. *The MDA algorithm outputs the minimum sum-rate $R_{ACO}(V)$, the fundamental partition \mathcal{P}^* and an optimal rate $\mathbf{r}_V \in \mathcal{R}_{ACO}(V)$. The estimation of $R_{ACO}(V)$, α , converges monotonically upward to $R_{ACO}(V)$.*

Example IV.2. *For the system in Example II.1, we start the MDA algorithm with singleton partition $\mathcal{P} = \{\{1\}, \dots, \{5\}\}$ and $\alpha = \sum_{i \in V} \frac{H(V) - H(\{i\})}{|V| - 1} = \frac{19}{4}$. Let the linear ordering be $\Phi = (4, 3, 2, 5, 1)$. By calling the CoordSatCapFus algorithm, we have the following results.*

We initiate $\mathbf{r}_V = (\alpha - H(V))\chi_V = (-\frac{13}{4}, \dots, -\frac{13}{4})$ and set $\mathcal{P}^ = \{\{4\}\}$ and $r_4 = f_{19/4}^\#(\{4\}) = \frac{7}{4}$ so that $\mathbf{r}_V = (-\frac{13}{4}, -\frac{13}{4}, -\frac{13}{4}, \frac{7}{4}, -\frac{13}{4})$.*

- For $\phi_2 = 3$, the values of $f_\alpha^\#(\{\phi_2\} \cup \tilde{U}) - r(\{\phi_2\} \cup \tilde{U})$

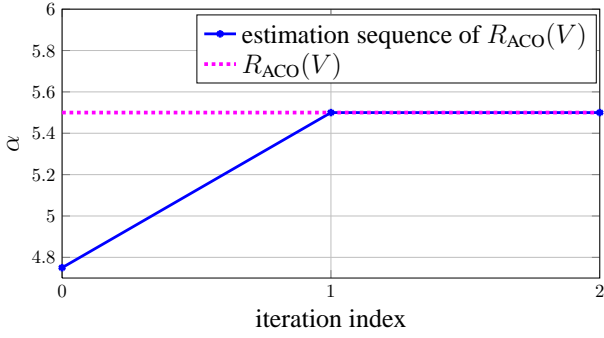


Fig. 1. The estimation sequence of $R_{ACO}(V)$, i.e., the value of α in each iteration, when the MDA algorithm is applied to the system in Example IV.2.

for all $U \subseteq \mathcal{P}^* = \{\{4\}\}$ are

$$f_{19/4}^\#(\{3\}) - r(\{3\}) = 6, f_{19/4}^\#(\{3, 4\}) - r(\{3, 4\}) = 21/4.$$

So, the saturation capacity $\hat{\xi} = 21/4$, the minimal minimizer $U^* = \{\{4\}\}$ and $U_4^* = \{\{3\}, \{4\}\}$. We update to $r_3 = -\frac{13}{4} + \frac{21}{4} = 2$ so that $\mathbf{r}_V = (-\frac{13}{4}, -\frac{13}{4}, 2, \frac{7}{4}, -\frac{13}{4})$. We have only one element $\{4\} \in \mathcal{P}^*$ such that $\tilde{U}_4^* \cap \{4\} \neq \emptyset$. So, $\mathcal{X} = \{\{4\}\}$ and $\tilde{U}_{\phi_i}^* \cup \mathcal{X} = \{3, 4\}$. We update to $\mathcal{P}^* = \{\{3, 4\}\}$.

- For $\phi_3 = 2$, the values of $f_\alpha^\#(\{\phi_3\} \cup \tilde{U}) - r(\{\phi_3\} \cup \tilde{U})$ for all $U \subseteq \mathcal{P}^* = \{\{3, 4\}\}$ are

$$\begin{aligned} f_{19/4}^\#(\{2\}) - r(\{2\}) &= 3, \\ f_{19/4}^\#(\{2, 3, 4\}) - r(\{2, 3, 4\}) &= 17/4. \end{aligned}$$

We have $\hat{\xi} = 3$ and $U_2^* = \{\{2\}\}$. We update to $\mathbf{r}_V = (-\frac{13}{4}, -\frac{1}{4}, 2, \frac{7}{4}, -\frac{13}{4})$. Since $\tilde{U}_2^* \cap X = \emptyset, \forall X \in \mathcal{P}$, we have $\mathcal{X} = \emptyset$ and $\mathcal{P}^* = \{\{3, 4\}, \{2\}\}$.

- For $\phi_4 = 5$, we have $\hat{\xi} = 3$, $U_5^* = \{\{5\}\}$ and $\mathcal{X} = \emptyset$. We update to $\mathbf{r}_V = (-\frac{13}{4}, -\frac{1}{4}, 2, \frac{7}{4}, -\frac{1}{4})$ and $\mathcal{P}^* = \{\{3, 4\}, \{2\}, \{5\}\}$.
- For $\phi_5 = 1$, we have $\hat{\xi} = \frac{13}{4}$, $U_1^* = \{\{3, 4\}, \{1\}\}$ and $\mathcal{X} = \{\{3, 4\}\}$. Therefore, the CoordSatCapFus algorithm terminates with $\mathbf{r}_V = (0, -\frac{1}{4}, 2, \frac{7}{4}, -\frac{1}{4})$ and $\mathcal{P}^* = \{\{1, 3, 4\}, \{2\}, \{5\}\}$.

Since $\mathcal{P} \neq \mathcal{P}^*$, we continue the iteration in the MDA algorithm. In the second iteration, we have $\mathcal{P} = \{\{1, 3, 4\}, \{2\}, \{5\}\}$ and $\alpha = \frac{11}{2}$. The CoordSatCapFus algorithm returns $\mathbf{r}_V = (0, \frac{1}{2}, 2, \frac{5}{2}, \frac{1}{2})$ and $\mathcal{P}^* = \{\{1, 3, 4\}, \{5\}, \{2\}\}$. The MDA algorithm terminates since $\mathcal{P} = \mathcal{P}^*$. One can show that the outputs \mathbf{r}_V , \mathcal{P}^* and α are respectively an optimal rate in $\mathcal{R}_{ACO}^*(V)$, the fundamental partition and the minimum sum-rate $R_{ACO}(V)$ for asymptotic model. We plot the value of α in each iteration, or the estimation sequence of $R_{ACO}(V)$, in Fig. 1. It can be shown that α converges monotonically upward to $R_{ACO}(V)$.

The CoordSatCap algorithm is one of the standard tools for solving the Dilworth truncation problem in the literature, e.g., [11]. It is also used in [6], [7] to determine an optimal rate vector in $\mathcal{R}_{NCO}^*(V)$ and/or checking whether a sum-rate α is

achievable for non-asymptotic setting.¹ But, in these works, the CoordSatCap algorithm is implemented on the original user set instead of a fused one. For example, in [6], [7], the the saturation capacity $\hat{\xi}$ is determined by the SFM problem

$$\min\{f_\alpha^\#(X) - r(X) \mid \phi_i \in X \subseteq V_i\}, \quad (6)$$

where $V_i = \{\phi_1, \dots, \phi_i\}$. Problem (6) can be solved in $O(\text{SFM}(|V_{i-1}|))$ time, where $\text{SFM}(|V|)$ denotes the complexity of an SFM algorithm for a set function defined on 2^V . On the contrary, the corresponding SFM problem

$$\min\{f_\alpha^\#(\{\phi_i\} \cup \tilde{U}) - r(\{\phi_i\} \cup \tilde{U}) : U \subseteq \mathcal{P}^*\} \quad (7)$$

in step 4 in the CoordSatCapFus algorithm is done over \mathcal{P}^* , a fused/merged user sets of V_{i-1} that is obtained by steps 8 and 9 in the previous iterations. Here, the objective function in (7) is submodular on $2^{\mathcal{P}^*}$. Problem (7) can be solved in $\text{SFM}(|\mathcal{P}^*|)$ time. Since $|\mathcal{P}^*| \leq |V_2|$, (7) is less complex than (6). For example, in the first iteration of the MDA algorithm when $\phi_3 = 2$ in Example IV.2, We have $\mathcal{P}^* = \{\{3, 4\}\}$ and $V_2 = \{3, 4\}$ such that $|\mathcal{P}^*| < |V_2|$. Problem (7) completes in $O(\text{SFM}(1))$ time, while problem (6) completes in $O(\text{SFM}(2))$ time.² See the experimental results in Section VI.

B. Non-asymptotic Model

The algorithms in [6], [7] for non-asymptotic CO model can adjust α on the nonnegative integer grid until it finally reaches $R_{NCO}(V)$, where the CoordSatCap can be replaced by the CoordSatCapFus algorithm which is less complex. See experimental results in Section VI.

In fact, the value of $R_{NCO}(V)$ and an optimal rate in $\mathcal{R}_{NCO}^*(V)$ can be determined by one more call of the CoordSatCapFus algorithm after solving the asymptotic CO problem. Let $R_{ACO}(V)$ be the asymptotic minimum sum-rate determined by the MDA algorithm. We know automatically $R_{NCO}(V) = \lceil R_{ACO}(V) \rceil$. By calling the CoordSatCapFus algorithm with input $\alpha = R_{NCO}(V)$, we can determine the value of an optimal rate in $B(\hat{f}_{R_{NCO}(V)}^\#, \leq) \cap \mathbb{Z}^{|V|} = \mathcal{R}_{NCO}^*(V)$. The integrality of this optimal vector is shown in Section V.

Example IV.3. Assume that we get $R_{ACO}(V) = \frac{11}{2}$ in Example IV.2. Then, $R_{NCO}(V) = \lceil R_{ACO}(V) \rceil = 6$. By calling

$$(\mathbf{r}_V, \mathcal{P}^*) \leftarrow \text{CoordSatCapFus}(V, H, R_{NCO}(V), \Phi),$$

we have $\mathbf{r}_V = (0, 1, 2, 3, 0)$ for linear ordering $\Phi = (4, 3, 2, 5, 1)$ and $\mathcal{P}^* = \{\{1, 2, 3, 4, 5\}\}$,³ where \mathbf{r}_V is an optimal rate in $\mathcal{R}_{NCO}^*(V)$ for non-asymptotic model.

¹If the sum-rate α is not achievable, the rate $\mathbf{r}_V \in B(\hat{f}_\alpha^\#, \leq)$ returned by the CoordSatCap algorithm has $r(V)$ strictly less than α .

²In the case when $|V| = 1$, SFM reduces to comparison between two possible sets, empty and ground sets, i.e., it is not necessary to call the SFM algorithm. This example just shows the difference in complexity.

³For $R_{NCO}(V) > R_{ACO}(V)$, the minimizer of $\min_{\mathcal{P} \in \Pi(V)} f_{R_{NCO}(V)}^\#[\mathcal{P}]$ is uniquely $\{V\}$ [19].

V. MINIMUM WEIGHTED SUM-RATE PROBLEM

Let $\mathbf{w}_V = (w_i : i \in V) \in \mathbb{R}_+^{|V|}$ and $\mathbf{w}_V^\top \mathbf{r}_V = \sum_{i \in V} w_i r_i$. We say that $\Phi = (\phi_1, \dots, \phi_{|V|})$ is a linear ordering that is consistent with \mathbf{w}_V if $w_{\phi_1} \leq w_{\phi_2} \leq \dots \leq w_{\phi_{|V|}}$.

Theorem V.1. *Let Φ be the linear ordering consistent with \mathbf{w}_V . The optimal rate \mathbf{r}_V returned by the MDA algorithm for asymptotic model is the minimizer of $\min\{\mathbf{w}_V^\top \mathbf{r}_V : \mathbf{r}_V \in \mathcal{R}_{\text{ACO}}^*(V)\}$; The optimal rate \mathbf{r}_V returned by CoordSatCapFus($V, H, \lceil R_{\text{ACO}}(V) \rceil, \Phi$) for asymptotic model is the minimizer of $\min\{\mathbf{w}_V^\top \mathbf{r}_V : \mathbf{r}_V \in \mathcal{R}_{\text{NCO}}^*(V)\}$.*

Proof: In the last iteration of the MDA algorithm, we call the CoordSatCapFus algorithm by inputting $\alpha = R_{\text{ACO}}(V)$. The Dilworth truncation $\hat{f}_{R_{\text{ACO}}(V)}^\#$ is a polymatroid rank function [3]. Let $\text{EX}(\hat{f}_{R_{\text{ACO}}(V)}^\#)$ be the set that contains all extreme points, or vertices, of the base polyhedron $B(\hat{f}_{R_{\text{ACO}}(V)}^\#, \leq)$. We have the initial point $\mathbf{r}_V = (\alpha - H(V))\chi_V \leq \mathbf{r}'_V, \forall \mathbf{r}'_V \in \text{EX}(\hat{f}_{R_{\text{ACO}}(V)}^\#)$.⁴ So, the CoordSatCapFus algorithm necessarily returns an extreme point in $B(\hat{f}_{R_{\text{ACO}}(V)}^\#, \leq)$ which minimizes $\min\{\mathbf{w}_V^\top \mathbf{r}_V : \mathbf{r}_V \in \mathcal{R}_{\text{ACO}}^*(V)\}$ [16]. In the same way, we can prove the claim for the non-asymptotic model. In addition, $f_{R_{\text{NCO}}(V)}^\#$ is integer-valued. So is $\hat{f}_{R_{\text{NCO}}(V)}^\#$. Therefore, all extreme points in $B(\hat{f}_{R_{\text{NCO}}(V)}^\#, \leq)$ are integral. ■

For example, one can show that $\mathbf{r}_V = (0, \frac{1}{2}, 2, \frac{5}{2}, \frac{1}{2})$ in Example IV.2 and $\mathbf{r}_V = (0, 1, 2, 3, 0)$ in Example IV.3 are the minimum weighted sum-rate vector in $\mathcal{R}_{\text{ACO}}^*(V)$ and $\mathcal{R}_{\text{NCO}}^*(V)$, respectively, where the weight \mathbf{w}_V is the one that linear ordering $\Phi = (4, 3, 2, 5, 1)$ is consistent with, e.g., $\mathbf{w}_V = (4, 0.5, 0.5, 0.3, 3.3)$.

Note, any linear ordering is consistent with $\mathbf{w}_V = (1, \dots, 1)$, i.e., if the problem is just to determine the minimum sum-rate and an optimal rate vector, the linear ordering can be arbitrarily chosen.

VI. COMPLEXITY

The authors in [10] proposed a divide-and-conquer (DV) algorithm for the asymptotic CO problem. The idea is to directly apply the DA algorithm in [11] to determine the fundamental partition and iteratively break each non-singleton subsets in it into singletons to determine each tuple in the optimal rate. Since the DA algorithm completes in $O(|V|^2 \cdot \text{SFM}(|V|))$ time, the complexity of the DV algorithm is upper bounded by $O(|V|^3 \cdot \text{SFM}(|V|))$. The complexity of the MDA algorithm is upper bounded by $O(|V|^2 \cdot \text{SFM}(|V|))$,⁵ which is lower than the DV algorithm.

Let $|V|$ be the size of the SFM problem with complexity $\text{SFM}(|V|)$. As aforementioned, although the numbers of calls of SFM algorithm are the same, the size of each SFM problem in the CoordSatCapFus algorithm based on (7) is less than that in the CoordSatCap algorithm based on (6) in general.

⁴ $\mathbf{r}_V \leq \mathbf{r}'_V, \forall \mathbf{r}'_V \in \text{EX}(f)$ is a tighter condition than $\mathbf{r}_V \in P(f, \leq)$.

⁵The complexity of the CoordSatCapFus algorithm based on (7) in the worst case is the same as the CoordSatCap algorithm based on (6). The worst case is when $\mathcal{P}^* = \{\{\phi_1\}, \dots, \{\phi_i\}\}$ for all i in the CoordSatCapFus algorithm. In the DA algorithm in [11], the CoordSatCap algorithm is implemented for solving the Dilworth truncation problem. Therefore the complexity of MDA algorithm is upper bounded by $O(|V|^2 \cdot \text{SFM}(|V|))$.

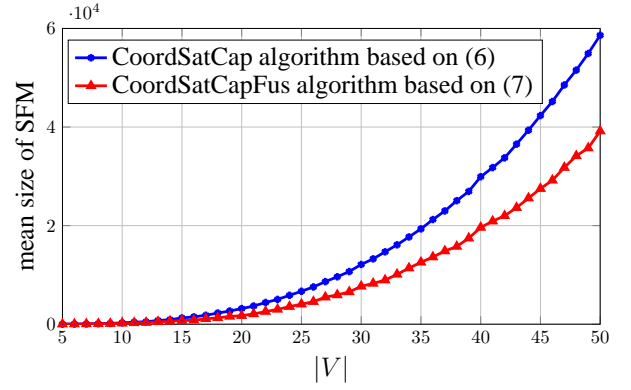


Fig. 2. The size of SFM problem over repetitions in the experiment in Section VI, where $H(V)$ is fixed to 50 and $|V|$ varies from 5 to 50.

We do an experiment to compare the complexity of these two algorithms. Let $H(V)$ be fixed to 50 and $|V|$ vary from 5 to 30. For each value of $|V|$, we repeat the procedure for 20 times: (a) randomly generate a CO system; (b) apply the MDA algorithm twice, one calls the CoordSatCapFus algorithm and the other calls the CoordSatCap algorithm. We record overall/summed size of the SFM algorithm in each run of the MDA algorithm and average over the repetitions. The results are shown in Fig. 2. It can be seen that by implementing the CoordSatCapFus algorithm, there is a considerable reduction in complexity when the size of user set $|V|$ grows.

VII. CONCLUSION

We proposed an MDA algorithm for determining the minimum sum-rate and a corresponding optimal rate for the asymptotic CO problem. The MDA algorithm mainly proposed an idea on how to update the minimum sum-rate estimation: A closer estimation to the optimum could be obtained by the minimal/finest minimizer of a Dilworth truncation problem based on the current estimation. We also proposed a CoordSatCapFus algorithm to solve the Dilworth truncation problem which was less complex than the original CoordSatCap algorithm. We discussed how to extend the MDA algorithm to solve the non-asymptotic problem and how to choose a proper linear ordering of the user set to solve a minimum weighted sum-rate problem.

APPENDIX A

PROOF OF THEOREM IV.1

In [11], [19], the authors proposed a DA for determining the principal partition sequence (PSP) for a clustering problem. Since the fundamental partition is one of the partitions in PSP [3], [18], we adapt DA to MDA to just determine the fundamental partition. A similar approach can be found in [18]. Based on the studies in [18], [19], if the CoordSatCapFus algorithm is able to determine the minimum and the minimal/finest minimizer of the Dilworth truncation problem $\min_{\mathcal{P} \in \Pi(V)} f_\alpha^\#[\mathcal{P}]$ for a given value of α , the MDA algorithm outputs $R_{\text{ACO}}(V)$, the fundamental partition and an optimal rate $\mathbf{r}_V \in \mathcal{R}_{\text{ACO}}^*(V) = B(\hat{f}_{R_{\text{ACO}}(V)}^\#, \leq)$. In addition, the value

of α of the MDA algorithm converges monotonically upward to $R_{\text{ACO}}(V)$.

Now, we show that CoordSatCapFus algorithm determines the finest minimizer of $\min_{\mathcal{P} \in \Pi(V)} f_{\alpha}^{\#}[\mathcal{P}]$. For $f_{\alpha}^{\#}$, Consider the (original/general) CoordSatCap algorithm [16]:

step 1: Initiate \mathbf{r}_V such that $\mathbf{r}_V \in P(f_{\alpha}^{\#}, \leq)$;
 step 2: For each dimension $i \in \{1, \dots, |V|\}$, do $\mathbf{r} \leftarrow \mathbf{r} + \hat{\xi} \chi_{\phi_i}$,
 where $\hat{\xi}$ is the saturation capacity

$$\hat{\xi} = \min\{f_{\alpha}^{\#}(X) - r(X) : \phi_i \in X \subseteq V\}. \quad (8)$$

$\hat{\xi}$ in (8) is the maximum increment in r_{ϕ_i} such that the resulting \mathbf{r}_V is still in $P(f_{\alpha}^{\#}, \leq)$, hence the name saturation capacity. Due to the intersecting submodularity of $f_{\alpha}^{\#}$, (8) is an SFM problem and the CoordSatCap algorithm finally updates \mathbf{r}_V to a vector/rate in $B(\hat{f}_{\alpha}^{\#}, \leq)$ with $r(V) = \hat{f}_{\alpha}^{\#}(V)$.

The minimal minimizer of $\min_{\mathcal{P} \in \Pi(V)} f_{\alpha}^{\#}[\mathcal{P}]$ is determined as follows. Let \hat{X}_{ϕ_i} be the minimal minimizer of (8) for dimension ϕ_i . By iteratively merging dimensions $\phi_i, \phi_j \in V$ such that $\phi_i \in \hat{X}_{\phi_j}$ until there is no such pair left, we can determine the finest partition in $\Pi(V)$ that minimizes $f_{\alpha}^{\#}[\mathcal{P}]$ [11], [16], [20].⁶ The implementation is as follows. Initiate $\mathcal{P}^* = \{\{\phi_i\} : i \in V\}$ at the beginning of the CoordSatCap algorithm. After obtaining each \hat{X}_{ϕ_i} for i in step 2, do the followings:

- find all elements in \mathcal{P}^* that intersect with \hat{X}_{ϕ_i} , i.e., determine $\mathcal{X} = \{X \in \mathcal{P}^* : X \cap \hat{X}_{\phi_i} \neq \emptyset\}$;
- merge all the elements in \mathcal{X} to form a single element in \mathcal{P}^* by $\mathcal{P}^* = (\mathcal{P}^* \setminus \mathcal{X}) \cup \hat{\mathcal{X}}$.

\mathcal{P}^* is the minimal minimizer of $\min_{\mathcal{P} \in \Pi(V)} f_{\alpha}^{\#}[\mathcal{P}]$ at the end of the CoordSatCap algorithm. It is easy to see that by letting $\mathbf{r}_V = (\alpha - H(V))\chi_V$ we have $\mathbf{r}_V \in P(f_{\alpha}^{\#}, \leq)$ initially. Let Φ be any linear ordering of V . We have

$$\begin{aligned} \min\{f_{\alpha}^{\#}(X) - r(X) : \phi_i \in X \subseteq V\} \\ = \min\{f_{\alpha}^{\#}(X) - r(X) : \phi_i \in X \subseteq V_i\} \end{aligned} \quad (9)$$

where $V_i = \{\phi_1, \dots, \phi_i\}$ due to the monotonicity of the entropy function H [16].⁷

Lemma A.1. *Let \mathcal{P}^* be the partition that is updated in each iteration of the CoordSatCap algorithm as described above,*

$$\begin{aligned} \min\{f_{\alpha}^{\#}(X) - r(X) : \phi_i \in X \subseteq V\} \\ = \min\{f_{\alpha}^{\#}(\tilde{U}) - r(\tilde{U}) : \{\phi_i\} \in U \subseteq \mathcal{P}^*\}. \end{aligned}$$

Let \hat{X}_{ϕ_i} and $U_{\phi_i}^$ be the minimal minimizer of the LHS and RHS, respectively, of the equation above. Then, $\hat{X}_i = \tilde{U}_i^*$.*

⁶The minimal minimizer of $\min_{\mathcal{P} \in \Pi(V)} f_{\alpha}^{\#}[\mathcal{P}]$ corresponds to the minimal separators of a submodular system with the rank function being $\hat{f}_{\alpha}^{\#}$. Define the partial order \preceq as $\phi_i \preceq \phi_j$ if $\phi_i \in \hat{X}_{\phi_j}$. Let $G(V, E)$ be the digraph with the edge set constituted by edges $e_{\phi_i, \phi_j} \in E$ if $\phi_i \preceq \phi_j$. The minimal separators are the strongly connected components of the underlying undirected graph of $G(V, E)$. The procedure that updates \mathcal{P}^* in Appendix A is exactly the one that determines these minimal separators. For more details, we refer the reader to [16], [20].

⁷This property has also been used in [6], [7] for solving the non-asymptotic CO problem.

Proof: For any $X \subseteq V$, let $\mathcal{Y} = \{Y \in \mathcal{P} : Y \cap X \neq \emptyset\}$. We have

$$\begin{aligned} f_{\alpha}^{\#}(X) - r(X) &= f_{\alpha}^{\#}(\tilde{\mathcal{Y}}) + r(\tilde{\mathcal{Y}}) \\ &= f_{\alpha}^{\#}(X) - f_{\alpha}^{\#}(\tilde{\mathcal{Y}}) + r(\tilde{\mathcal{Y}} \setminus X) \\ &= f_{\alpha}^{\#}(X) - f_{\alpha}^{\#}(\tilde{\mathcal{Y}}) + \sum_{Y \in \mathcal{Y}} (f_{\alpha}^{\#}(Y) - f_{\alpha}^{\#}(Y \cap X)) \geq 0, \end{aligned}$$

where the last inequality is obtained by applying submodular inequality (3) inductively over intersecting subsets. The minimality of \tilde{U}_i^* over all $X \subseteq V$ such that $\phi_i \in X$ can also be seen by induction. So, $\hat{X}_i = \tilde{U}_i^*$. ■

Based on (9) and Lemma A.1, we can implement the CoordSatCap algorithm by a fusion method as in the CoordSatCapFus algorithm, where steps 8 and 9 are equivalent to the procedure that updates \mathcal{P}^* as described above.

REFERENCES

- [1] I. Csiszár and P. Narayan, "Secrecy capacities for multiple terminals," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3047–3061, Dec. 2004.
- [2] S. El Rouayheb, A. Sprintson, and P. Sadeghi, "On coding for cooperative data exchange," in *Proc. IEEE Inf. Theory Workshop*, Cairo, 2010, pp. 1–5.
- [3] C. Chan, A. Al-Bashabsheh, J. Ebrahimi, T. Kaced, and T. Liu, "Multivariate mutual information inspired by secret-key agreement," *Proc. IEEE*, vol. 103, no. 10, pp. 1883–1913, Oct. 2015.
- [4] C. Chan, A. Al-Bashabsheh, J. B. Ebrahimi, T. Kaced, S. Kadhe, T. Liu, A. Sprintson, M. Yan, and Q. Zhou, "Successive omniscience," in *Proc. Int. Symp. Network Coding*, Sydney, 2015, pp. 21–25.
- [5] C. Chan, Al-Bashabsheh, Q. Zhou, N. Ding, T. Liu, and A. Sprintson, "Successive omniscience," *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 3270–3289, Apr. 2016.
- [6] N. Milosavljevic, S. Pawar, S. E. Rouayheb, M. Gastpar, and K. Ramchandran, "Efficient algorithms for the data exchange problem," *IEEE Trans. Inf. Theory*, vol. 62, no. 4, pp. 1878 – 1896, Feb. 2015.
- [7] T. Courtade and R. Wesel, "Coded cooperative data exchange in multi-hop networks," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 1136–1158, Feb. 2014.
- [8] N. Ding, R. A. Kennedy, and P. Sadeghi, "Fairest constant sum-rate transmission for cooperative data exchange: An M -convex minimization approach," in *Proc. 22nd Int. Conf. Telecommun.*, Sydney, Australia, 2015, pp. 36–42.
- [9] —, "Estimating minimum sum-rate for cooperative data exchange," in *Proc. IEEE Int. Symp. Inf. Theory*, Hong Kong, China, 2015, pp. 2618–2622.
- [10] N. Milosavljevic, S. Pawar, S. El Rouayheb, M. Gastpar, and K. Ramchandran, "Deterministic algorithm for the cooperative data exchange problem," in *Proc. IEEE Int. Symp. Inf. Theory*, St. Petersburg, 2011, pp. 410–414.
- [11] K. Nagano, Y. Kawahara, and S. Iwata, "Minimum average cost clustering," in *Proc. Advances in Neural Inf. Process. Syst.*, Vancouver, 2010, pp. 1759–1767.
- [12] R. W. Yeung, *Information theory and network coding*. Berlin, Germany: Springer Science & Business Media, 2008.
- [13] N. Ding, C. Chan, T. Liu, R. A. Kennedy, and P. Sadeghi, "A game-theoretic perspective on communication for omniscience," in *Proc. 2016 Australian Commun. Theory Workshop*, Melbourne, Australia, 2016, pp. 95–100.
- [14] N. Ding, R. A. Kennedy, and P. Sadeghi, "Iterative merging algorithm for cooperative data exchange," in *Proc. Int. Symp. Network Coding*, Sydney, Australia, 2015, pp. 41–45.
- [15] S. Fujishige, "Polymatroidal dependence structure of a set of random variables," *Inf. and Control*, vol. 39, no. 1, pp. 55 – 72, Oct. 1978.
- [16] —, *Submodular functions and optimization*, 2nd ed. Amsterdam, The Netherlands: Elsevier, 2005.
- [17] R. P. Dilworth, "Dependence relations in a semi-modular lattice," *Duke Math. J.*, vol. 11, no. 3, pp. 575–587, 1944.
- [18] C. Chan, A. Al-Bashabsheh, Q. Zhou, T. Kaced, and T. Liu, "Info-clustering: A mathematical theory for data clustering," *arXiv preprint arXiv:1605.01233*, 2016.

- [19] H. Narayanan, “The principal lattice of partitions of a submodular function,” *Linear Algebra and its Appl.*, vol. 144, pp. 179 – 216, Jan. 1991.
- [20] R. E. Bixby, W. H. Cunningham, and D. M. Topkis, “The partial order of a polymatroid extreme point,” *Math. Oper. Res.*, vol. 10, no. 3, pp. 367–378, 1985.